知识管理论坛　ISSN 2095-5472　CN11-6036/C
Knowledge Management Forum

E-mail: kmf@mail.las.ac.cn
http://www.kmf.ac.cn

【学术探索】

基于文献计量学定律的大数据应用领域发展规律研究

◎ 张姣姣　刘云　程旖婕

北京理工大学管理与经济学院 北京 100081

摘要：［目的／意义］研究大数据应用领域的研究现状和发展趋势，并揭示其发展规律。［方法／过程］制定大数据应用技术的数据检索策略，选取1990-2015年相关文献研究数据，借助计量软件VP（Vantage Point）、统计分析软件SPSS和Microsoft Excel等工具，分别从载文量分布、期刊分布和作者分布3个层面，验证并分析该领域文献发展是否符合普赖斯科学文献增长规律、布拉德福定律以及洛特卡定律。［结果／结论］大数据应用领域相关论文自1990年开始，经过一段时间的稳步发展后，至2012年开始显现突飞猛进的发展，其文献发展符合普赖斯科学文献增长规律；文献样本的期刊分布基本符合布拉德福定律，并已形成核心期刊群，其中包括BMC BIOINFORMATICS、SENSORS等；该领域著者分布与洛特卡定律相差很大，尚未形成较具影响力的核心著者群。

关键词：大数据应用　普赖斯科学文献增长规律　布拉德福定律　洛特卡定律

分类号：G250

引用格式：张姣姣，刘云，程旖婕．基于文献计量学定律的大数据应用领域发展规律研究[J/OL]. 知识管理论坛，2016, 1(5): 384-392[引用日期]. http://www.kmf.ac.cn/p/1/50/.

1 引言

近年来，“大数据”引起各国政府、产业界和科技界的高度关注。全球范围内，运用大数据推动经济发展、完善社会治理、提升政府服务和监管能力正成为趋势，有关发达国家相继制定实施大数据战略性文件，大力推动大数据发展和应用[1]。联合国“数据脉动”计划[2]、美国“大数据”战略[3]、日本“面向2020年的ICT综合战略”[4]、韩国大数据中心战略[5]等先后开启了大数据战略的大幕。大量跨国企业也进入大数据研究与开发领域，如传统数据分析企业天睿公司（Teradata）、思爱普（SAP）、SPSS以及大数据资源企业谷歌、Facebook等[4]。2008年，Nature杂志出版专刊Big Data，基于多个学科的实际研究现状系统地介绍了“大数据”所蕴含的潜在价值与挑战[6]；2011年，Science杂志出版的专刊Dealing with Data，这些研究标志着大

基金项目：本文系国家自然科学基金项目“我国专利申请量增长的影响因素及动力机制研究”（项目编号:71273030）和“国家创新体系国际化政策协同机制/过程模型及效应评估研究”（项目编号:71573017）研究成果之一。

作者简介：张姣姣（ORCID：0000-0001-8281-710X），硕士研究生，E-mail:zhangjjcx@163.com；刘云（ORCID:0000-0002-2888-8932），教授，博士生导师；程旖婕（ORCID：0000-0002-9365-055X），博士研究生。

收稿日期：2016-08-01　　发表日期：2016-10-27　　本文责任编辑：徐健

知识管理论坛，2016(5):384–392
DOI：10.13266/j.issn.2095–5472.2016.044

数据”正式登上科学研究的大舞台，并成为各个学科中的研究热点问题。一个大规模生产、分享和应用数据的时代即将到来 [8]。

大数据应用（big data application，BDA）是运用大数据的思维和方法，利用大数据分析的结果，为用户提供辅助决策，发掘潜在价值的过程 [9]，越来越多的问题可以通过大数据的应用来解决。其应用不仅仅限于科学技术层面，在公共管理、基础研究和应用研究以及商业领域等层面，大数据都能够带来新理念和新思维，提供解决问题的新方法或新视角 [8]。高小平 [10]、李欢 [11]、高霞 [12] 分别在国家治理变革创新、科技管理创新平台构建和科技评价方法等方面，运用了大数据的思维和方法；从基础和应用研究角度，李振皓 [14] 和黄少芳 [15] 分别在中药的质量控制和地质资料档案信息化的相关研究中应用了大数据，N. O. E. Olsson 和 H. Bullberg [16] 在项目评价的过程中结合了对大数据的应用，S. R. Sukumar 和 R. Natarajan [17] 则将大数据方法应用到了卫生保健领域的相关研究中，M. C. Ebach 和 M. S. Michael [18] 等人曾研究了历史科学和大数据的关系；此外，在商业领域，T. H. Davenport [19], J. Frizzo-Barker [20] 和 C. F. Hofacker [21] 等运用大数据分别从国际商业决策、商业奖学金和消费者行为等不同角度进行了系列研究。可见“大数据”已经在很多领域得到广泛应用，在本质上已经转化为一种新的思维方式、一种新的问题解决方法 [22]。

大数据应用的益处是多角度多方面的，不仅能够优化政府效能、管理决策、市场监管、公共服务、城市设施、社会治安等，还可以为经济、教育、文化、卫生、外交等领域带来不菲的收益。我国互联网、移动互联网用户规模居全球第一，拥有丰富的数据资源和应用市场优势，大数据部分关键技术研发取得突破，涌现出一批互联网创新企业和创新应用，一些地方政府已启动大数据相关工作 [1]。充分利用我国的数据规模优势，实现数据规模、质量和应用水平同步提升，发掘和释放数据资源的潜在价值，有利于更好发挥数据资源的战略作用。因此，无论是科学技术层面、公共管理部门、基础和应用研究、商业领域等层面的大数据应用，还是经济、教育、文化、卫生、外交等领域的应用研究，都具有较高的研究意义 [8]。

本研究着眼于大数据应用，探索该领域的文献发展规律，继而进行相关发展定律的验证与研究，以期系统地揭示 BDA 技术的发展特征、规律和趋势，从而为后续相关研究提供参考。

2 研究方法与数据来源

关于大数据概念的起源说法不一，有人认为最早是由全球知名咨询公司麦肯锡提出的，也有人认为是 1998 年由 SGI 的首席科学家 J. R. Mashey 在一篇发表于 USENIX 会议的名为“Big Data and the Next Wave of Infrastres”的文章中首次提出的 [3]；还有人认为是由 20 世纪 90 年代数据仓库之父 B. Inmon 提出的。本研究基于 Web of Science 数据库，制定了相关数据检索策略，将所有涉及 BDA 概念或运用大数据思维或方法来解决有关问题的论文均作为研究分析的数据对象，检索结果中的数据包含 1990-2015 年的数据信息，共计 1 701 项。研究使用的数据库 Web of Science，是一种综合性文献检索工具，其数据来源于世界上 40 多个国家和地区中的学术期刊、专题文集和会议录，以及大量的专利文献和科技图书等。涉及的学科包括生物学、农学、医学、化学、物理学、地球科学、生命科学等，是国际上收集科学论文相互引证最为完备的数据库。本研究选取科学引文索引（SCI）为基础数据源，制定相应检索策略，在此基础之上进行进一步的研究和分析。

科技文献自产生以来便随时间的推移而增长，普赖斯通过考察科技文献的积累过程，从中发现了增长规律。这一成果被收集在他的代表著作《巴比伦以来的科学》中，对后世的科学文献研究产生了深刻的影响 [23]。普赖斯对摘要性杂志的增长率进行了推算，以科技文献量为纵轴，以历史年代为横轴，把不同年代的科

技文献量在坐标图上逐点描绘出来，然后用一条光滑曲线联结各点，十分近似的表征了科技文献随时间增长的规律，这就是著名的普赖斯曲线，其数学表达式为：

$$F(t)=ae^{bt}(a>0,b>0) \quad (1)$$

其中 F(t) 表示科技文献量，a 为统计初始时刻（即 t=0）的文献量，e 为自然对数底，b 是期刊的持续增长率，是一个时间常数。

布拉德福定律和洛特卡定律都属于文献计量学的重要定律，与齐普夫定律一起被并称为文献计量学的三大定律。布拉德福定律可以表述为：如果将科技期刊按其刊载某学科专业论文的数量多少，以递减顺序排列，那么可以把期刊分为专门面向这个学科的核心区、相关区和非相关区。各个区的文章数量相等，此时核心区、相关区、非相关区期刊数量呈的关系。布拉德福定律是以科学论文在期刊中的分散情况为基础创建的，但却可以派生出许多不同的应用，对于确定核心期刊、制定文献采购策略、优化馆藏、检验工作情况、了解读者阅读倾向、检索利用文献等方面都有一定的指导作用 [24]。

洛特卡定律是美国人洛特卡在 1926 年发现的。该定律被认为是第一次揭示了作者频率与文献数量之间的关系，描述科学生产率的频率分布规律。广义的洛特卡定律可以用以下公式表示：

$$f(x)=cn^{-n} \quad (2)$$

其中：f(x) 表示在一定时间内，某一特定学科或主题领域中，撰写了 x 篇论文的作者数（或作者频率）；c 和 n 是两个大于零的常数 [25]。一般对洛特卡定律的描述为：发表两篇论文的作者数量约为发表一篇论文的作者数量的 1/4；发表三篇论文的作者数量约为发表一篇论文作者数量的 1/9；发表 N 篇论文的作者数量约为发表一篇论文作者数量的 $1/n^2$ ……，而发表一篇论文作者的数量约占所有作者数量的 60%（即取 c=1，n=2）。本研究即采用这种方法进行分析和验证。

在采用以上 3 种文献计量学方法的同时，本研究还借助了计量软件 VP（Vantage Point）、统计分析软件 SPSS 和 Microsoft Excel 等工具，对所获取数据进行载文、期刊和著者情况等层面的统计、分析和验证。

3 统计结果分析

3.1 载文量分析

文献发表数量从一定程度上反映出该领域的相关研究在既定时间的研究水平和发展趋势。文献的数量随时间、地区、类型等变量分布，是文献样本的基本特征之一，也是最基础、最简单的分析项目之一 [26]。本研究统计了 1990-2015 年这 16 年间 BDA 领域的文献统计量的变化情况，见图 1。

从所收录相关文献的时间来看，最早的文献是 1990 年来自匈牙利布达佩斯科技经济大学的科学家 T. Kantor 所著的题为 *Novel Applications of halogenation Reactions in Atomic Spectrometry* 的文章。他将大数据应用到原子光谱分析卤化反应的研究当中，是早期应用大数据进行科学研究的良好范例。根据所获得的时间 - 文献量二维数据表，可以进一步得到随时间变化的载文量的趋势和载文量增长率的变化情况（见图 1）：①从 1990 年的第一篇文献收录开始，1991- 997 年的 7 年间，文献量均维持在 10 篇左右，大数据的应用技术研究尚处在初期发展期间；② 1998- 2011 年，文献量较上一阶段有一定增长，除个别年份有较高增长率和负增长外，基本维持较低水平增长率，可将 1990-2011 年界定为 BDA 的初步稳定发展期；③ 2012-2015 年文献量大幅增加，2015 年收录的文献量已经达到 509 篇，与初期发展阶段（1990-2011 年）的文献总量 603 篇相差无几；这 4 年间文献量增长率也提升至 50% 左右，从文献量的折线图来看，基本呈现直线上升趋势，增长势头较猛，2014 年甚至达到 92.02%，4 年年均增长率达 66.27%，可见该时期为 BDA 的迅速发展阶段。

为了更加科学合理地掌握 BDA 技术领域相

知识管理论坛，2016(5):384-392
DOI：10.13266/j.issn.2095-5472.2016.044

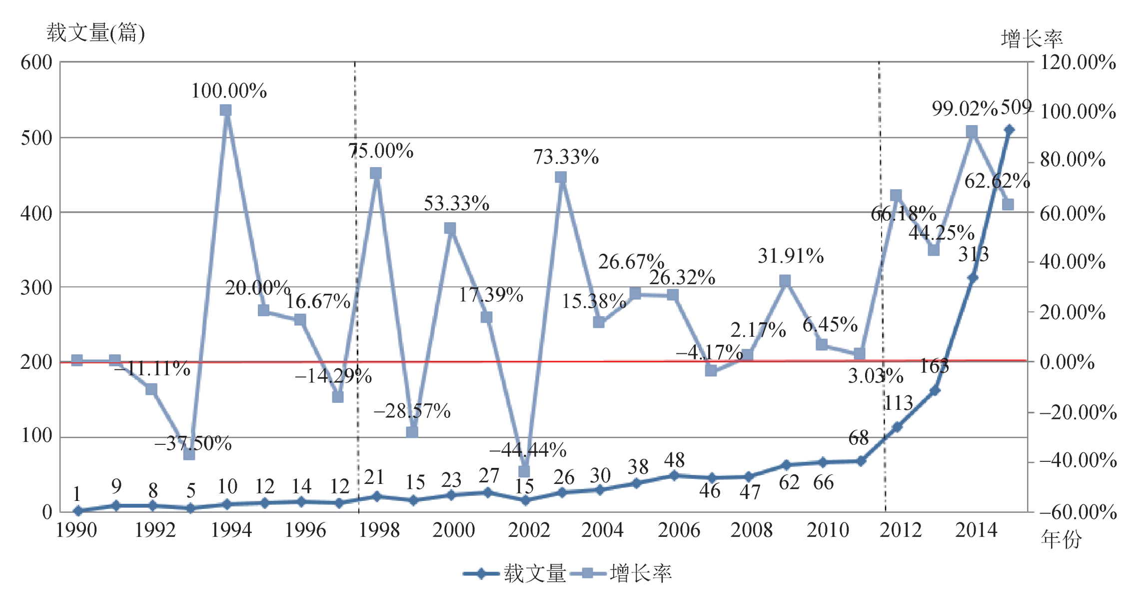


图 1 载文量变化及其增长率

关文献研究的发展规律，本研究对该领域文献研究的增长规律进行了时间序列的统计研究，研究过程借助 SPSS 统计分析软件对该领域的文献增长情况进行曲线拟合，以验证其发展规律是否符合普莱斯科学文献增长规律。本研究以时间（年份）为自变量，以累计文献量为因变量，在 SPSS 统计分析软件中对相关统计数据进行了曲线拟合。其中包括线性拟合、二次曲线拟合、三次曲线拟合、对数函数拟合、复合函数拟合、增长曲线拟合以及指数拟合。根据拟合结果显示，其中线性、对数函数拟合模型的相关系数 R^2 较小（分别为 0.714 和 0.440），予以排除，其余相关系数较大的模型包括二次曲线拟合模型、三次曲线拟合模型、复合函数模型、增长函数模型以及指数模型，依据普赖斯文献增长规律中刻画的普赖斯曲线表达式。选择指数拟合模型来计算 BDA 领域文献量的拟合方程，可以得到如下表达式：

$$F(t)=8.909\times0.205^{t}=8.909\times0.205^{t} \quad (3)$$

本模型中 t（年份）的初始值为 1，即 1990 年设置为 1，之后依次递增，该拟合模型的拟合分析结果（模型拟合情况表、方差分析表以及系数表）见表 1，曲线拟合图见图 2。对指数函数模型拟合的结果分析如下：

（1）模型的拟合情况反映了模型对数据的解释能力。修正的可决系数（调整 R^2）越大，模型的解释能力越强。对表 1 的结果分析可知，本模型的 R^2 为 0.882，说明该模型对数据的解释能力较好。

表 1 模型拟合情况 (模型汇总)

R	R^2	调整	估计值的标准误差
0.942	0.887	0.882	0.572

（2）方差分析反映了模型整体的显著性，一般将模型的检验 P 值（Sig.）与 0.05 作比较，如果小于 0.05，即为显著。对表 2 的结果分析可知，该模型的显著性水平为 0.000，小于 0.05，说明该模型是显著的。

表 2 方差分析 (ANOVA)

项目	平方和	df	均方	F	Sig.
回归	61.590	1	61.590	188.104	0.000
残差	7.858	24	0.327	-	-
总计	69.449	25	-	-	-

（3）回归系数是各个变量在回归方程中的系数值，Sig. 值表示回归系数的显著性，越



小越显著。对表 2 的结果分析可知，该模型中，常数项是 8.909，年份的系数是 0.157，t 值是 13.715，显著性是 0.000。可见两个结果都是显著的。

表 3 系数统计

项目	未标准化系数		标准化系数	t	Sig.
	B	标准误差	Beta		
年份	0.205	0.015	0.942	13.715	0.000
常数	8.909	2.059	-	4.328	0.000

（4）图 2 是线性、二次函数、三次函数和指数函数的拟合情况，图中的圆圈表示实际值，不难看出，指数函数拟合模型的拟合效果是最好的。由表 1- 表 3 分析可知，该模型的值最高，整体 P 值最小，即拟合优度最好，显著性最好。

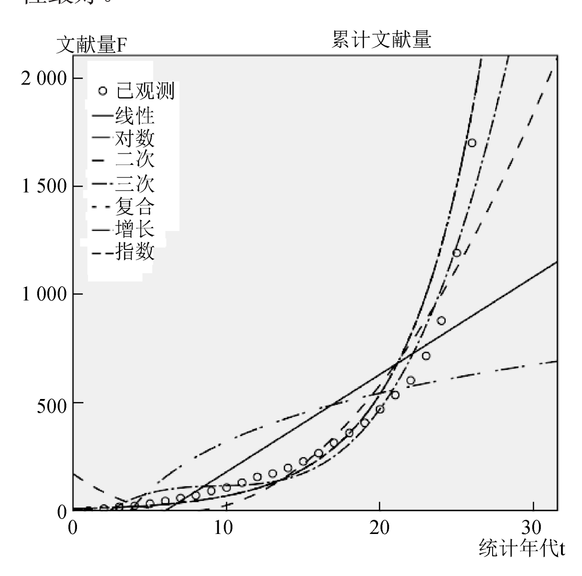


图 2 拟合曲线

由以上结果分析可以得到以下相关结论 :BDA 领域的文献量处于指数增长期，即迅速发展的阶段；由此可以推测，在未来一段时间内，该领域的相关研究还会继续快速增长，文献量也将呈指数增长（一定时期内）。综合以上分析结果可以认为，BDA 领域的文献发展符合普赖斯科学文献增长规律。

3.2 期刊分析

英国著名文献计量学家布拉德福最早提出了核心期刊的概念。虽然布拉德福的定义完全是按照期刊发表论文数量而确定的，具有一定局限性，而现在对于核心期刊的定义更多的考虑了期刊的使用率（包括被引率、文摘率、流通率）和学术影响力等因素，但布拉德福文献集中分散定律一直是核心期刊评价的理论基础[27]。本文采用布拉德福定律的分析方法对 BDA 领域文章的期刊分布规律进行整理和研究。

所选取数据源中的文献来自于 886 种不同期刊，其中来自期刊 *Future Generation Computer Systems-The International Journal of Grid Computing and Escience* 的文章最多，为 30 篇；刊载文章超过 10 篇（含 10 篇）的期刊有 14 种。具体分析如下：

根据布拉德福定律中关于核心区、相关区和非相关区期刊数量的比例关系，本研究将 886 种期刊的载文量分成大致 3 个区域，其中第一区域约占总数 6.4% 的期刊（57 种期刊）发表了总量 30% 的文献，第二区域约占总数 21.3% 的期刊（189 种期刊）发表了总量 30% 左右的文献，第三区域约占总数 72.2% 的期刊（640 种期刊）发表了总量 40% 左右的文献，表明 BDA 领域的文献呈现明显的核心密集分布。3 个区域期刊的分布数量比为 57:189:640，这与布拉德福定律中的 $1:n:n^2$ 相拟合，其中 n 约为 3.3。由以上数据和分析可以判断，本研究文献样本的期刊分布基本符合布拉德福定律。从 1990 年时只有 1 本期刊发表 1 篇文章，发展至 2015 年累计共有 886 种期刊共发表 1 701 篇文章，BDA 领域已经形成广泛的研究覆盖范围并形成自己的核心期刊群。具体如表 4、表 5、图 3 所示：

表 4 期刊分布列表

对比项	文献量（篇）	文献量占比	期刊数量(种)	百分比
核心区	510	29.98%	57	6.43
相关区	512	30.10%	189	21.33
非相关区	679	39.92%	640	72.23
合计	1 701	100.00%	886	100.00

知识管理论坛，2016(5):384–392
DOI：10.13266/j.issn.2095–5472.2016.044

表 5 期刊名称列表（载文量≥ 10）

序号	文献量（篇）	累计文献量	百分比	累计百分比	期刊名
1	30	30	1.76	1.76	*Future Generation Computer Systems-The International Journal of Grid Computing and Escience*
2	24	54	1.41	3.17	*International Journal of Distributed Sensor Networks*
3	23	77	1.35	4.53	*BMC Bioinformatics*
4	20	97	1.18	5.70	*PLOS ONE*
5	16	113	0.94	6.64	*Big Data*
6	16	129	0.94	7.58	*Sensors*
7	14	143	0.82	8.41	*Cluster Computing-The Journal of Networks Software Tools and Applications*
8	11	154	0.65	9.05	*IBM Journal of Research and Development*
9	11	165	0.65	9.70	*IEEE Access*
10	11	176	0.65	10.35	*IEEE Transactions on Parallel and Distributed Systems*
11	11	187	0.65	10.99	*Journal of Parallel and Distributed Computing*
12	10	197	0.59	11.58	*IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*
13	10	207	0.59	12.17	*Journal of Internet Technology*
14	10	217	0.59	12.76	*Transportation Research Record*

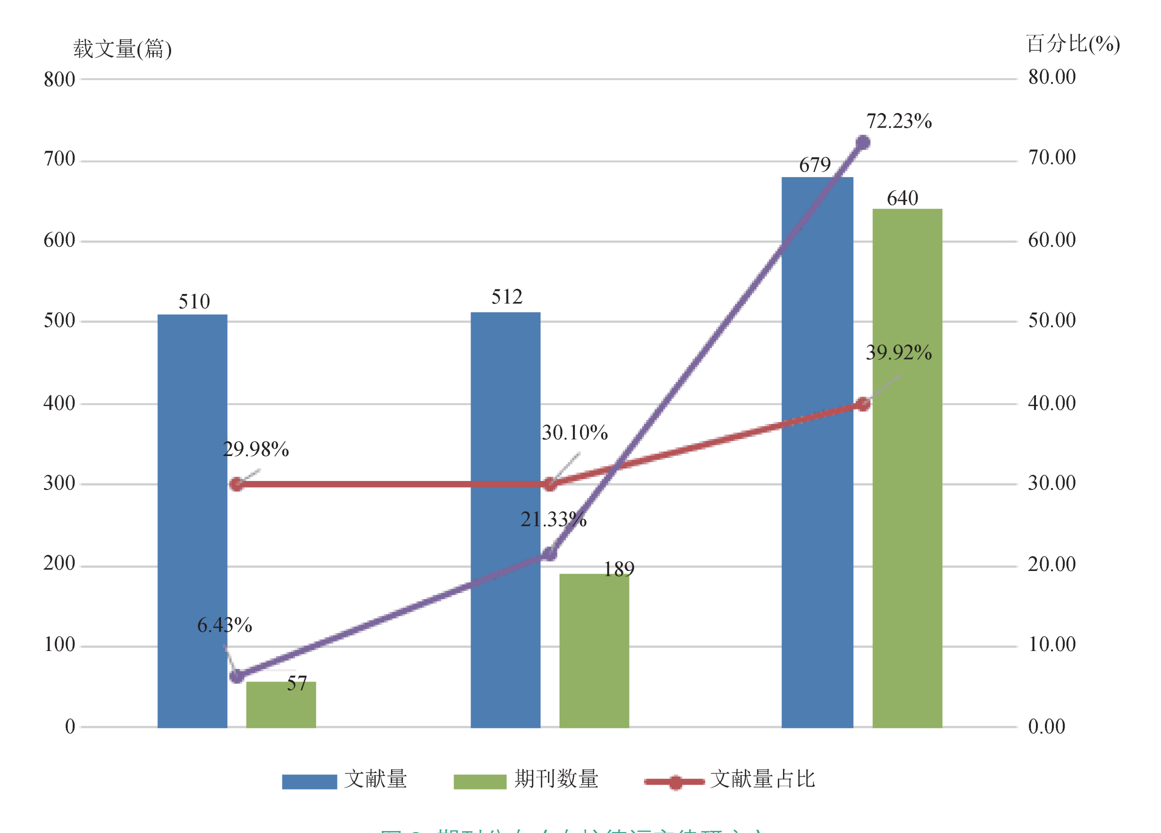


图 3 期刊分布（布拉德福定律研究）

3.3 著者分布规律

著者是论文的重要外部特征之一，是决定论文质量的关键[23]。核心著者的评定需要综合考虑多方面的因素和指标，洛特卡定律则只从发文量角度评定某一领域的核心著者，虽然这种评定方式忽略了对著者发文的‘质’的考量，同时忽视了在同一篇文章中不同著者的贡献程度等因素，但运用洛特卡定律来刻画某领域文献及著者的分布情况在一定程度上是不失公道的。

研究统计 1990-2015 年在 BDA 领域发表文献的实际著者情况，并计算出按照洛特卡定律所估计的著者数，在此基础之上比较二者的相对误差。具体情况如图 4 所示：

图 4 文献实际著者分布与洛特卡定律分布比较

从图 4 可以看出，在高发文量的著者群中，其著者数过少，如发文量为 12 篇、11 篇和 9 篇的著者均只有 1 人。而按照洛特卡定律估计来看，其文献著者数量远不止这些，分别应有 26 人、31 人和 47 人。可见该领域的实际著者分布与按照洛特卡定律估计的著者分布相差较多。但发文 1 篇的作者为 5 916 人，洛特卡定律估计数量为 3 809，实际情况远远超出估计。综上所述，BDA 领域的文献著者分布与按照洛特卡定律估计的著者分布相差很大；从整体的分布情况来看，在一定程度上反映出大数据应用的发展尚未达到成熟稳定的阶段。

4 研究结论及启示

研究通过对普赖斯科学文献增长规律、布拉德福定律和洛特卡定律的验证，分别从载文、期刊和著者 3 个层面，对 BDA 领域的文献增长情况进行了分析和研究。根据研究结果，可以得出以下几方面的结论及启示：

(1) 根据对载文的分析结果可以观察到，自 1990 年起，BDA 领域的相关文献经过一段时间的稳步发展后，直至 2012 年开始显现突飞

知识管理论坛，2016(5):384−392
DOI：10.13266/j.issn.2095−5472.2016.044

猛进的发展态势，该领域文献整体呈指数增长态势，发展符合普赖斯科学文献增长规律。采用此验证方法不仅可以较为简单、准确地描绘BDA领域过去的发展态势，还可据此预测其未来几年将继续沿此态势发展，可以更好地刻画其发展轨迹和整体走向。

(2) 从期刊角度来看，BDA 领域已形成核心期刊群，其中包括 *BMC Bioinformatics*、*Sensors* 等，该期刊群在一定程度上代表着该领域的发展前沿，此外，据分析结果显示，文献样本的期刊分布基本符合布拉德福定律。依据此结论，可以针对核心期刊群的期刊及其所刊载的文章进行深入研究，以更加及时、准确地掌握该领域发展的前沿动态。

(3) 从著者分布情况来看，该领域著者分布与洛特卡定律相差很大，尚未形成较具影响力的核心著者群，在一定程度上反映出该领域的相关研究尚未成熟。此外，还可以认为，对于一个发展尚未成熟、尚处于发展探索阶段的新兴领域，要想形成一个较权威较具影响力的核心研究群，是需要时间和经验的积累。

(4) 本研究应用经典的计量学定律，研究和验证了 BDA 领域的发展规律。以此类推，对于其他的新兴领域的研究，也可以运用这种验证经典定律的方法，以更好地掌握该新兴领域的发展趋势和状态。

从近年来大数据技术的迅猛发展来看，无论是各国政府、国际组织、社会企业，还是高校机构、各学科领域，都给予了极大的关注和良好预期，而最重要的环节当属对于大数据技术在各行各业的应用。虽然在 BDA 技术上存在一定的挑战，但相信这更是一次良好的推动人类社会整体进步的机遇，大数据应用是大势所趋，是社会进步的新方向。本文分别从载文、期刊和著者层面揭示了 BDA 领域的现状和发展趋势，希望能够在一定程度上揭示该领域的研究规律并预测未来的发展趋势，以期为后续相关研究提供有价值的参考。

研究过程中，由于条件有限，有许多因素尚未纳入考虑范围，这可能在一定程度上导致研究的不准确和不全面。后续笔者将加强全方位的考虑，以更加准确和全面的掌握 BDA 领域的发展情况。

参考文献：

[1] 国务院关于印发促进大数据发展行动纲要的通知[EB/OL]. [2016-04-21]. http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm.
[2] 联合国“全球脉动”计划发布《大数据开发： 机遇与挑　战 》[EB/OL]. [2016-04-23]. http://www.docin.com/p-750680124.html.
[3] MASHEY J R. Big data and the next wave of infra stress[C]//Computer Science Division Seminar. Berkeley: University of California, Berkeley, 1997.
[4] 苏红 . 大数据对 H 公司科技创新战略的影响 [D]. 青岛 : 中国海洋大学 , 2014.
[5] 罗梓超 , 吕志坚 . 韩国大数据发展及应用分析 [J]. 全球科技经济瞭望 , 2014(3): 22-26.
[6] Nature Specials Archive. Big data[EB/OL]. [2016-04-29]. http://www.nature.com/news/specials/bigdata/index.html.
[7] Science, Special Online Collection. Dealing with data [EB/OL]. [2016-02-11]. http://www.sciencemag.org/site/special/data/.
[8] 大数据时代降临 [EB/OL]. [2016-05-02]. http://www.banyuetan.org/jrt/120922/70953_1.shtml.
[9] 张引 , 陈敏 , 廖小飞 . 大数据应用的现状与展望 [J]. 计算机研究与发展 , 2013, 50(S1): 216-233.
[10] 高小平 . 借助大数据科技力量寻求国家治理变革创新 [J]. 中国行政管理 , 2015(10): 10-14.
[11] 李欢 . 大数据背景下科技管理创新平台构建研究 [J]. 科学管理研究 , 2014, 32(3): 44-48.
[12] 高霞 . 基于大数据的科技评价研究 [J]. 科技预测与评价 , 2015(11): 27-30.
[13] 崔燕 . 大数据时代的专利业务数据质量评价探析 [J]. 中国发明与专利 , 2013(9): 68-71.
[14] 李振皓 , 钱忠直 , 程翼宇 . 基于大数据科技的中药质量控制技术创新战略 [J]. 中国中药杂志 , 2015, 40(17): 3374-3378.
[15] 黄少芳 , 刘晓鸿 . 基于大数据的地质资料档案信息化与服务 [J]. 资源与产业 , 2015, 17(6): 56-61.
[16] OLSSON N O E, BULLBERG H. Use of big data in project evaluations [J]. International journal of managing projects in business, 2015, 8(3): 491-512.
[17] SUKUMAR S R, NATARAJAN R, FERRELL R K. Quality of big data in health care [J]. International journal of health care quality assurance, 2015, 28(6): 621-634.
[18] EBACH M C, MICHAEL M S, SHAW W S, et al. Big data and the historical sciences: a critique [J]. Geoforum,

2016, 71: 1-4.
[19] DAVENPORT T H. How strategists use “big data” to support internal business decisions, discovery and production[J]. Strategy & leadership, 2014, 42(4): 45-50.
[20] FRIZZO-BARKER J, CHOW-WHITE P A, MOZAFARI M, et al. An empirical study of the rise of big data in business scholarship[J]. International journal of information management, 2016, 36(3): 403-413
[21] HOFACKER C F, MALTHOUSE E C, SULTAN F. Big data and consumer behavior: imminent opportunities[J]. Journal of consumer marketing, 2016, 33(2): 89-97.
[22] 金陵 . 大数据与信息化教学变革 [J]. 中国电化教育 , 2013(10): 8-13.
[23] 赵双 . 普莱斯对科学计量学的贡献及当代意义 [C]. 第三届全国科技哲学暨交叉学科研究生论坛 . 北京 : 中国自然辩证法研究会 , 2010.
[24] 邱均平 . 信息计量学 (四) 第四讲 文献信息离散分布规律 —— 布拉德福定律 [J]. 情报理论与实践 , 2000(4): 315-320.
[25] 王静 , 王宏鑫 . 关于 Lotka 定律的研究——纪念洛特卡定律创立 80 周年 [J]. 情报杂志 , 2007, 26(4): 94-96.
[26] 傅余洋子 , 华薇娜 . 基于 Web of Science 数据库中云计算研究文献的计量分析 [J]. 新世纪图书馆 , 2013(7): 57-64.
[27] 刘雪立 . 中文核心期刊评价指标体系 : 演进 . 问题 . 建议 [J]. 编辑学报 , 2014, 26(1): 92-95.

作者贡献说明：
张姣姣：设计研究框架，开展研究工作，撰写论文；
刘　云：协同构思和设计研究框架，修改论文；
程漪婕：收集和分析数据，协助撰写论文。

Research on Developing Regulations of Big Data Application Technology Based on Bibliometrics Laws

Zhang Jiaojiao　Liu Yun　Cheng Yijie
School of Management and Economics, Beijing Institute of Technology, Beijing 100081

Abstract: [Purpose/significance] This paper aims to study the trend over the past and the present situation of big data application, and indicate the development regulation and trend in the future. **[Method/process]** This paper analyzed literature year distribution, journal distribution and author distribution of big data application field, and was the first to verify whether or not its development corresponds to the three fundamental bibliometric laws, though there are lots of factors that have not been taken into consideration. **[Result/conclusion]** Results shows that starting in 1990, literatures related to big data application field went through a period of stable development, and appears to develop rapidly from 2012, whose development corresponds to Price law of scientific literature growth. Development of the literature sample in the study is in line with Bradford’s law, and forms a group of core journals, including Bmc bioinformatics, Sensors and so on. In the view of the author distribution, distribution in this filed is far different from that in Lotka’s law, and there’s no doubt that it has not yet formed the core author group.

Keywords: big data application　Price law of scientific literature growth　Bradford’s law　Lotka’s law